

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:27 ; Search time 91.75 Seconds

(without alignments)
15.135 Million cell updates/sec

Title: US-09-331-631a-1_COPY_74_116

Perfect score: 248
Sequence: 1 NOEDPQTECQCQRCRCRQEE.....RQOXYCQRCKEICEEEERY 43

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	42.3	588	1	VCLB_GOSHI
2	94	37.9	605	1	VCLB_GOSHI
3	78	31.5	47	1	AGRP_LUCFY
4	67	27.0	154	1	YS51_CAEEL
5	65	26.2	919	1	ANDR_HUMAN
6	64	25.8	911	1	ANDR_PANTR
7	63.5	25.6	33	1	MBP1_MAIZE
8	63.5	25.6	50	1	HSP1_MOUSE
9	63.5	25.6	50	1	HSP1_RAT
10	62	25.0	648	1	KAPC_DICDI
11	61	24.6	907	1	ANDR_CANFA
12	60	24.2	223	1	CAS2_SHEEP
13	59	23.8	285	1	INNO_CANFA
14	59	23.8	445	1	NRH3_MOUSE
15	59	23.8	447	1	NRH3_HUMAN
16	57.5	23.2	338	1	FSH_HUMAN
17	57	23.0	223	1	CAS2_CAPI
18	57	23.0	429	1	APPA_MACEA
19	57	23.0	577	1	MTG8_MOUSE
20	57	23.0	604	1	MTG8_HUMAN
21	57	23.0	758	1	YMS8_YEAST
22	57	23.0	1166	1	XMRK_XIPMA
23	57	23.0	1339	1	ERB3_RAT
24	57	23.0	2124	1	Y192_HUMAN
25	57	23.0	2318	1	MTG3_MOUSE
26	56.5	22.8	284	1	TPM1_RAT
27	56.5	22.8	284	1	TPM2_HUMAN
28	56.5	22.8	284	1	TPMB_HUMAN
29	56.5	22.8	284	1	TPMB_RABIT
30	56.5	22.8	1407	1	TRHY_RABIT
31	56	22.6	210	1	YOEI_ECOLI
32	56	22.6	524	1	SBP_SOYBN
33	56	22.6	626	1	AHI2_ARAHY

34	56	22.6	644	1	BTB_DROME	Q24266 drosophila
35	56	22.6	1210	1	EGFR_MOUSE	Q01279 mus musculu
36	56	22.6	1898	1	TRHY_HUMAN	Q07283 homo sapien
37	55.5	22.4	49	1	HSP1_PIG	P04101 sus scrofa
38	55.5	22.4	284	1	TPM1_CHICK	P04265 gallus gall
39	55.5	22.4	284	1	TPM1_MOUSE	P46901 mus musculu
40	55.5	22.4	284	1	TPMB_CHICK	P19352 gallus gall
41	55.5	22.4	303	1	BNK_DROME	P40794 drosophila
42	55.5	22.4	639	1	GLCX_SOYBN	P11827 glycine max
43	55.5	22.4	1391	1	MST2_DROHY	Q08696 drosophila
44	55	22.2	239	1	CALD_MERGA	P13505 melagris g
45	55	22.2	272	1	BMH2_YEAST	P34730 saccharomyc

ALIGNMENTS

```

RESULT 1
VCLB_GOSHI
ID VCLB_GOSHI STANDARD; PRT; 588 AA.
AC P09801;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RT "developmental biochemistry of cottonseed embryogenesis and
RT germination. XVIII. cDNA and amino acid sequences of the members of
RT the storage protein families.";
RL Plant Mol. Biol. 7:475-489(1986).
CC -i- FUNCTION: SEED STORAGE PROTEIN.
CC -i- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -i- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICTILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16891; AAA3071.1; -.
DR PIR: A30838; FMCNAB.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001113; -.
DR PRAM: PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1
FT CHAIN 25
FT SEQUENCE 588 AA; 69729 MW; 63E69B29A8ADEB CRC64;
SQ

```

Query Match 42.3%; Score 105; DB 1; Length 588;
Best Local Similarity 47.5%; Pred. No. 0.00035;
Matches 19; Conservative 9; Mismatches 10; Indels 2; Gaps 1;

```

OY 3 EDPTQECQRCRCRQEE--RQOXYCQRCKEICEEEERY 42
      ||| :||: ||||| ||: ||||: |:-|:
Db 81 EDPTQRECCQRCRCRQEE--RQRPCCQRCRLKRFQEQO 118

```

RESULT 2
VCLB_GOSHI
ID VCLB_GOSHI STANDARD; PRT; 605 AA.

```
AC P09799; DR 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Boritto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed.";
RL Plant Mol. Biol. 9:533-546(1987).
CC -I- FUNCTION: SEED STORAGE PROTEIN.
CC -I- SUBCELLULAR LOCATION: CYTOSOL; MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -.
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL
FT 1 23
FT CHAIN
FT 24 605 VICILIN GC72-A.
SQ SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;

Query Match 37.9%; Score 94; DB 1; Length 605;
Best Local Similarity 35.9%; Pred. NO. 0.0048;
Matches 14; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

QY 2 QEDPTQCQCQRRCROQESGPRQOQYCCRCKEICEE 40
ID 1: 1 : ::|||:: || 1 :: 1 : 1 : 1 :
DB 115 QQQPDKQFKECQQRCQWQEQRRERKQCVCKECRQYQGD 153

RESULT 3
AGRP_LURCY STANDARD: PRT: 47 AA.
AC P56568;
AD AGRP_LURCY ID AGRP_LURCY STANDARD: PRT: 47 AA.
AE 15-DEC-1998 (Rel. 37, Created)
AF 15-DEC-1998 (Rel. 37, Last sequence update)
AT 15-DEC-1998 (Rel. 37, Last annotation update)
AU 6.5 kDa ARGININE/Glutamate-RICH POLYPEPTIDE (6.5k-AGRP).
CA Biosci. Biotechnol. Biochem. 61:984-988(1997).
CB -I- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON
CD RESERVES DURING GERMINATION AND SEEDLING GROWTH.
CE -I- MASS SPECTROMETRY: MW=5693.39; METHOD=MALDI.
CF
```

```
CC -I- SIMILARITY: SOME, TO 7S SEED STORAGE PROTEINS.
KW Seed storage protein.
FT DISULFID 12 33
FT DISULFID 16 29
SQ SEQUENCE 47 AA; 5698 MW; 588B0E82273AC05 CRC64;

Query Match 31.5%; Score 78; DB 1; Length 47;
Best Local Similarity 40.0%; Pred. No. 0.028;
Matches 14; Conservative 9; Mismatches 8; Indels 4; Gaps 1;

OY 5 POTECCCCRCRQCESGPRQQGYCQRKCEICEE 39
   :|: |: |: |: |: |: ||||::|:|
DB 5 PTREYEACRVRCQVAHEGVYER-----QRRQGYCKE 35

RESULT 4
YS5L_CAEEL
ID YS5L_CAEEL STANDARD; PRT; 154 AA.
AC 009662;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHEETICAL 15.6 KDA PROTEIN ZK673.1 IN CHROMOSOME II PRECURSOR.
GN ZK673.1..
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabdilitis.
RN [1]
RP STRAIN-BRISTOL N2;
RA McMurtry A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP STRAIN-BRISTOL N2;
RA Jones S.J.M.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC -I- SIMILARITY: STRONG, TO C.ELEGANS E04D5.4 AND SOME, TO C.ELEGANS
CC C03G6.13.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC CC
CC DR EMBL; 248585; CAA8486.1; -.
CC DR HSSP; P01511; IACX.
CC DR WORMPEP; ZK673.1; CE16742.
CC KM Hypothetical protein; Signal.
CC FT SIGNAL 1 19 POTENTIAL.
CC FT CHAIN 20 154 HYPOTHEETICAL PROTEIN ZK673.1.
CC SQ SEQUENCE 154 AA; 15637 MW; 333A02C3A2E782AB CRC64;

Query Match 27.0%; Score 67; DB 1; Length 154;
Best Local Similarity 35.5%; Pred. No. 0.95;
Matches 11; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

OY 3 EDPTQCOCORCRQCESGPRQQGYCQRKCEICEE 33
   :|: |: |: |: |: |: ||||::|:|
DB 69 DDPNTRCTGYTSLCSNAKYTPLLQGFCRKTC 99

RESULT 5
ANDR_HUMAN
ID ANDR_HUMAN STANDARD; PRT; 919 AA.
AC P10275;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
```

DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR).
 GN AR OR NR3C4 OR DHTR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89112208; Pubmed=3216866;
 RA Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Larson R.E.,
 RA French F.S., Wilson E.M.;
 RT "The human androgen receptor: complementary deoxyribonucleic acid
 RT cloning, sequence analysis and gene expression in prostate.";
 RL Mol. Endocrinol. 2:1265-1275(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90083302; Pubmed=2594783;
 RA Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J.,
 RA Wilson E.M., French F.S.;
 RT "Sequence of the Intron/exon junctions of the coding region of the
 RT human androgen receptor gene and identification of a point mutation
 RT in a family with complete androgen insensitivity.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90258935; Pubmed=2342476;
 RA Govindan M.V.;
 RT "Specific region in hormone binding domain is essential for hormone
 RT binding and trans-activation by human androgen receptor.";
 RL Mol. Endocrinol. 4:417-427(1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX TISSUE=PROSTATE;
 RX MEDLINE=89017168; Pubmed=3174628;
 RA Chang C., Kokontis J., Lao S.;
 RT "Structural analysis of complementary DNA and amino acid sequences of
 RT human and rat androgen receptors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX TISSUE=PROSTATE;
 RX MEDLINE=89089609; Pubmed=2911578;
 RA Tilley W.D., Marcelli M., Wilson J.D., McPhaul M.J.;
 RT "Characterization and expression of a cDNA encoding the human androgen
 RT receptor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:327-331(1989).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX TISSUE=PROSTATE;
 RX MEDLINE=91155943; Pubmed=2293020;
 RA Marcelli M., Tilley W.D., Wilson C.M., Griffin J.E., Wilson J.D.,
 RA McPhaul M.J.;
 RT "Definition of the human androgen receptor gene structure permits the
 RT identification of mutations that cause androgen resistance: premature
 RT termination of the receptor protein at amino acid residue 588 causes
 RT complete androgen resistance.";
 RL Mol. Endocrinol. 4:1105-1116(1990).
 RN [7]
 RP SEQUENCE OF 189-919 FROM N.A.
 RX MEDLINE=88178111; Pubmed=3353726;
 RA Chang C., Kokontis J., Lao S.;
 RT "Molecular cloning of human and rat complementary DNA encoding
 RT androgen receptors.";
 RL Science 240:324-326(1988).
 RN [8]
 RP SEQUENCE OF 468-919 FROM N.A.
 RX MEDLINE=88240407; Pubmed=3377788;
 RA Trapman J., Klaassen P., Kuiper G.G.J.M., van der Korput J.A.G.M.,
 RA Faber P.W., van Rooij H.C.J., Geurts van Kessel A., Voorhorst M.M.,
 RA Mulder E., Brinkmann A.O.;
 RT "Cloning, structure and expression of a cDNA encoding the human
 RT androgen receptor.";
 RL Biochem. Biophys. Res. Commun. 153:241-248(1988).

RN [9]
 RP POLYMORPHISM OF POLY-GLN REGION.
 RX MEDLINE=92220629; Pubmed=1561105;
 RA Sledzins H.F., Oostra B.A., Brinkmann A.O., Trapman J.;
 RT "Trinucleotide repeat polymorphism in the androgen receptor gene
 RT (AR).";
 RL Nucleic Acids Res. 20:1427-1427(1992).
 RN [10]
 RP POLYMORPHISM OF POLY-GLY REGION.
 RX TISSUE=BLOOD.
 RC Lu J., Danielsen M.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN [11]
 RP VARIANTS SBMA IN POLY-GLN REGION.
 RX MEDLINE=91287825; Pubmed=2062380;
 RA la Spada A.R., Wilson E.M., Lubahn D.B., Harding A.E., Fischback K.H.;
 RT "Androgen receptor gene mutations in X-linked spinal and bulbar
 RT muscular atrophy.";
 RL Nature 352:77-79(1991).
 RN [12]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=95023089; Pubmed=7937057;
 RA Patterson M.N., Hughes L.A., Gottlieb B., Pinsky L.;
 RT "The androgen receptor gene mutations database.";
 RL Nucleic Acids Res. 22:3560-3562(1994).
 RN [13]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97169385; Pubmed=9016528;
 RA Gottlieb B., Trifiro M., Lumbroso R., Vasilou D.M., Pinsky L.;
 RT "The androgen receptor gene mutations database.";
 RL Nucleic Acids Res. 25:158-162(1997).
 RN [14]
 RP VARIANT LNCAP ALA-877.
 RX MEDLINE=91083633; Pubmed=2260966;
 RA Veldscholte J., Ris-Stalpers C., Kuiper G.G., Jenster G.,
 RA Berrevoets C., Klaassen E., van Rooij H.C.J., Trapman J.,
 RA Brinkmann A.O., Mulder E.;
 RT "A mutation in the ligand binding domain of the androgen receptor of
 RT human LNCAP cells affects steroid binding characteristics and
 RT response to anti-androgens.";
 RL Biochem. Biophys. Res. Commun. 173:534-540(1990).
 RN [15]
 RP VARIANT CAIS MET-866.
 RX MEDLINE=91186983; Pubmed=2082179;
 RA Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J.,
 RA Corfen J.L.;
 RT "Functional characterization of naturally occurring mutant androgen
 RT receptors from subjects with complete androgen insensitivity.";
 RL Mol. Endocrinol. 4:1759-1772(1990).
 RN [16]
 RP VARIANT CYS-774.
 RX MEDLINE=91310758; Pubmed=1856263;
 RA Marcelli M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,
 RA McPhaul M.J.;
 RT "Androgen resistance associated with a mutation of the androgen
 RT receptor at amino acid 772 (Arg-->Cys) results from a combination of
 RT decreased messenger ribonucleic acid levels and impairment of
 RT receptor function.";
 RL J. Clin. Endocrinol. Metab. 73:318-325(1991).
 RN [17]
 RP VARIANT CAIS ASN-695 AND HIS-695, AND SEQUENCE OF 629-723 FROM N.A.
 RX MEDLINE=92131007; Pubmed=1775137;
 RA Ris-Stalpers C., Trifiro M.A., Kuiper G.G., Jenster G., Romalo G.,
 RA Sai T., van Rooij H.C., Kaufman M., Rosenfield R.L., Lao S.;
 RT "Substitution of aspartic acid-686 by histidine or asparagine in the
 RT human androgen receptor leads to a functionally inactive protein with
 RT altered hormone-binding characteristics.";
 RL Mol. Endocrinol. 5:1562-1569(1991).
 RN [18]
 RP VARIANTS CAIS AND PAIS.
 RX MEDLINE=93338440; Pubmed=1307250;
 RA Batch J.A., Williams D.M., Davies H.R., Brown B.D., Evans B.A.J.,
 RA Hughes I.A., Patterson M.N.;

RA Duvick J.P., Rood T., Rao A.G., Marshak D.R.;
 RT "Purification and characterization of a novel antimicrobial peptide
 from maize (Zea mays L.) kernels.";
 RL J. Biol. Chem. 267:18814-18820(1992).
 CC -1- FUNCTION: INHIBITOR OF BOTH BACTERIAL AND FUNGAL GROWTH IN VITRO.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN THE EMBRYO PORTION OF THE
 CC KERNEL.
 DR PIR: A41822; A41822.
 DR MAZEDB: 69182; -.
 KW Fungicide; Antibiotic.
 SO SEQUENCE 33 AA; 4131 MW; B148F1B90E823599 CRC64;

Query Match 25.6%; Score 63.5; DB 1; Length 33;
 Best Local Similarity 44.0%; Pred. No. 0.64;
 Matches 11; Conservative 6; Mismatches 7; Indels 1; Gaps 1;

QY 11 OCORRC-RQESGPRQOYCQRCK 34
 Db 6 ECRROCLRHEGQWETQECMRCCR 30

RESULT 8
 HSPL_MOUSE STANDARD; PRT; 50 AA.
 AC P02319;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE SPERM PROTAMINE P1 (CYSTEINE-RICH PROTAMINE).
 GN PM1 OR PM-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85199803; PubMed=2986684;
 RA Kleene K.C., Distel R.J., Hecht N.B.;
 RT "Nucleotide sequence of a cDNA clone encoding mouse protamine 1.";
 RL Biochemistry 24:719-722(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88193085; PubMed=3358932;
 RA Johnson P.A., Paschon J.J., Yelick P.C., Palminter R.D., Hecht N.B.;
 RT "Sequence homologies in the mouse protamine 1 and 2 genes.";
 RL Biochim. Biophys. Acta 950:45-53(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87260978; PubMed=3037541;
 RA Paschon J.J., Behringer R.R., Brinster R.L., Palminter R.D.;
 RT "Spermatid-specific expression of protamine 1 in transgenic mice.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:5316-5319(1987).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88181903; PubMed=3445973;
 RA Hecht N.B.;
 RT "Gene expression during spermatogenesis.";
 RL Ann. N.Y. Acad. Sci. 513:90-101(1987).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX STRAIN=C129;
 RA Schluter G., Engel W.;
 RT Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
 CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
 CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
 CC -1- SUBUNIT: CROSS-LINKED BY INTERCHAIN DISULFIDE BONDS AROUND THE
 CC DNA-HELIX (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: TESTIS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; K02926; AAA39980.1; -.
 DR EMBL; X07625; CAA30472.1; -.
 DR EMBL; X14003; CAA32169.1; -.
 DR EMBL; M27500; AAA39985.1; -.
 DR EMBL; Z47352; CAA87410.1; -.
 DR PIR; A02660; HSMSS1.
 DR PIR; S03820; S03820.
 DR PIR; A28331; A28331.
 DR MGD; MGI:97765; PM1.
 DR INTERPRO; IPR000221; -.
 DR PFAM; PF00260; Protamine_P1; 1.
 DR PROSITE; PS00048; PROTAMINE_P1; 1.
 KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
 KW Testis; DNA condensation; Nuclear protein.
 FT INT.MET 0 0
 FT DISULFID 5 5 INTERCHAIN (WITH CYS-21) (BY SIMILARITY).
 FT DISULFID 6 14 BY SIMILARITY.
 FT DISULFID 21 21 INTERCHAIN (WITH CYS-5) (BY SIMILARITY).
 FT DISULFID 36 36 INTERCHAIN (WITH CYS-36) (BY SIMILARITY).
 FT DISULFID 37 47 BY SIMILARITY.
 SO SEQUENCE 50 AA; 6827 MW; B61CAF3D1BFF4978 CRC64;

Query Match 25.6%; Score 63.5; DB 1; Length 50;
 Best Local Similarity 28.9%; Pred. No. 0.88;
 Matches 11; Conservative 11; Mismatches 11; Indels 5; Gaps 1;

QY 6 OTECOCCORRCRQESGPRQOYCQRCKICEEERY 43
 Db 11 RSRCCRRRRRCR-----RRRCCRRRRRCRRRRSY 43

RESULT 9
 HSPL_RAT STANDARD; PRT; 50 AA.
 AC P10118;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE SPERM PROTAMINE P1 (CYSTEINE-RICH PROTAMINE).
 GN PM1 OR PM-1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89335257; PubMed=2757789;
 RA Klemm U., Lee C.H., Burfeind P., Hake S., Engel W.;
 RT "Nucleotide sequence of a cDNA encoding rat protamine and the haploid
 RT expression of the gene during rat spermatogenesis.";
 RL Biol. Chem. Hoppe-Seyler 370:293-301(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96341725; PubMed=8720108;
 RA Schluter G., Celik A., Obara R., Schlicker M., Hofferbert S.,
 RA Schlung A., Adam I.M., Engel W.;
 RT "Sequence analysis of the conserved protamine gene cluster shows that
 RT it contains a fourth expressed gene.";
 RL Mol. Reprod. Dev. 43:1-6(1996).
 RN [3]
 RP SEQUENCE.
 RX MEDLINE=89207111; PubMed=3072011;
 RA Ammer H., Henschel A.;
 RT "Rat sperm protamine. Isolation and sequence analysis.";
 RL Biol. Chem. Hoppe-Seyler 369:1301-1306(1988).
 RN [4]
 RP SEQUENCE OF 1-15 AND 44-50.


```
CC CC -I- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC at the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC DR EMBL; X03238; KAS26983.1; -.
CC CC DR PIR; A25070; KASHS2.
CC CC DR INTERPRO; IPR001588; -.
CC CC DR PFAM; PF00363; caseins; 2.
CC CC DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
CC CC KW Milk; Phosphorylation; Signal.
CC CC FT SIGNAL 1 15
CC CC FT CHAIN 16 223 ALPHA-S2 CASEIN.
CC CC FT REPEAT 77 141
CC CC FT REPEAT 159 223
CC CC FT MOD_RES 23
CC CC FT MOD_RES 24 23 PHOSPHORYLATION (POTENTIAL).
CC CC FT MOD_RES 25 25 PHOSPHORYLATION (POTENTIAL).
CC CC FT MOD_RES 72 72 PHOSPHORYLATION (POTENTIAL).
CC CC FT MOD_RES 73 73 PHOSPHORYLATION (POTENTIAL).
CC CC FT MOD_RES 74 74 PHOSPHORYLATION (POTENTIAL).
CC CC FT VARIANT 64 64 D->N
CC CC SO SEQUENCE 223 AA; 26332 MW; 67212935E27A26D7 CRC64;
CC CC -----
OY 1 NQDPQRTECQCQRCRRGDSPPRQQTCQCRKCTCE--EEET 43
DB 24 SSEEPIINSDEIVRKQNMALHPREKRKLCTSCSEEVANADEEY 68
CC CC -----
RESULT 13
ID INVO_CANFA STANDARD; PRT; 285 AA.
AC P18174;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DE INVOLUCRIN.
OS IVL.
GN Canis familiaris (dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90348475; PubMed=2385171;
RA Tseng H., Green H..
RT "The involucrin genes of pig and dog: comparison of their segments of
FT repeats with those of primates and higher primates.";
RL Mol. Biol. Evol. 7:293-302(1990).
CC CC -I- FUNCTION: INVOLUCRIN IS A KERATINOCTE PROTEIN THAT FIRST APPEARS
CC CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO THE
CC CC MEMBRANE PROTEINS BY TRANSGUTAMINASE. ALL THAT RESULTS IN THE
CC CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC CC -I- TISSUE SPECIFICITY: PRESENT IN KERATINOCTES OF EPIDERMIS AND
CC CC OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC CC -----
CC CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
CC CC AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE HTTP://WWW.ISB-SIB.CH/ANNOUNCE/
CC CC OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH).
```

